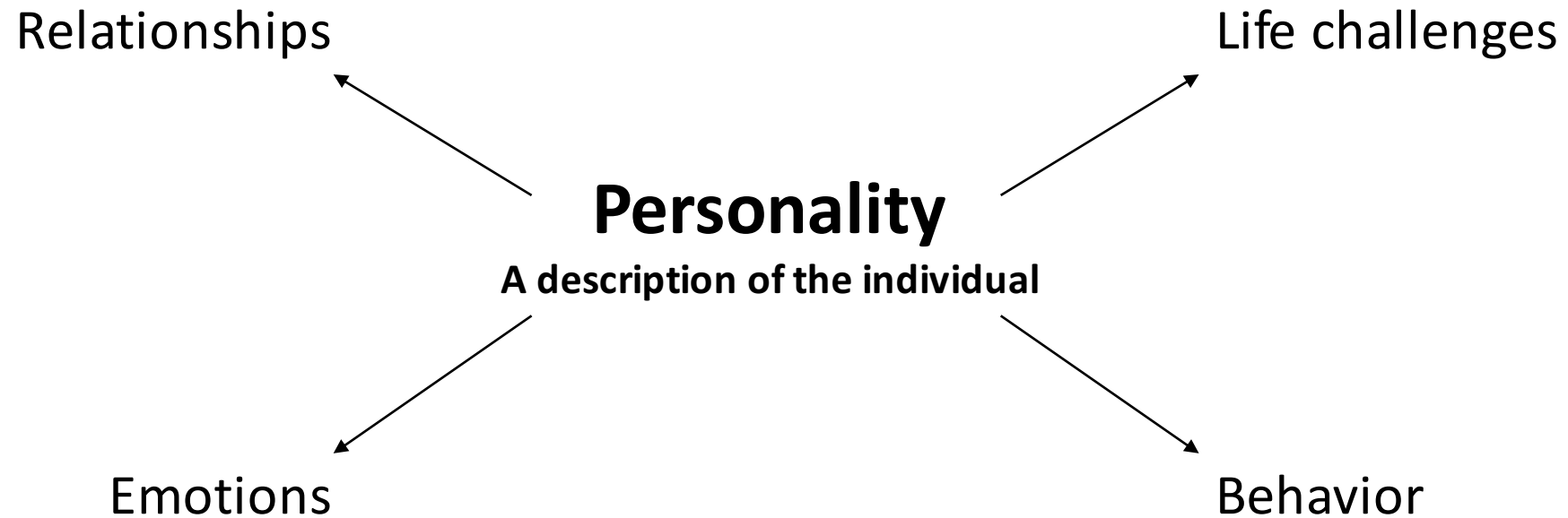


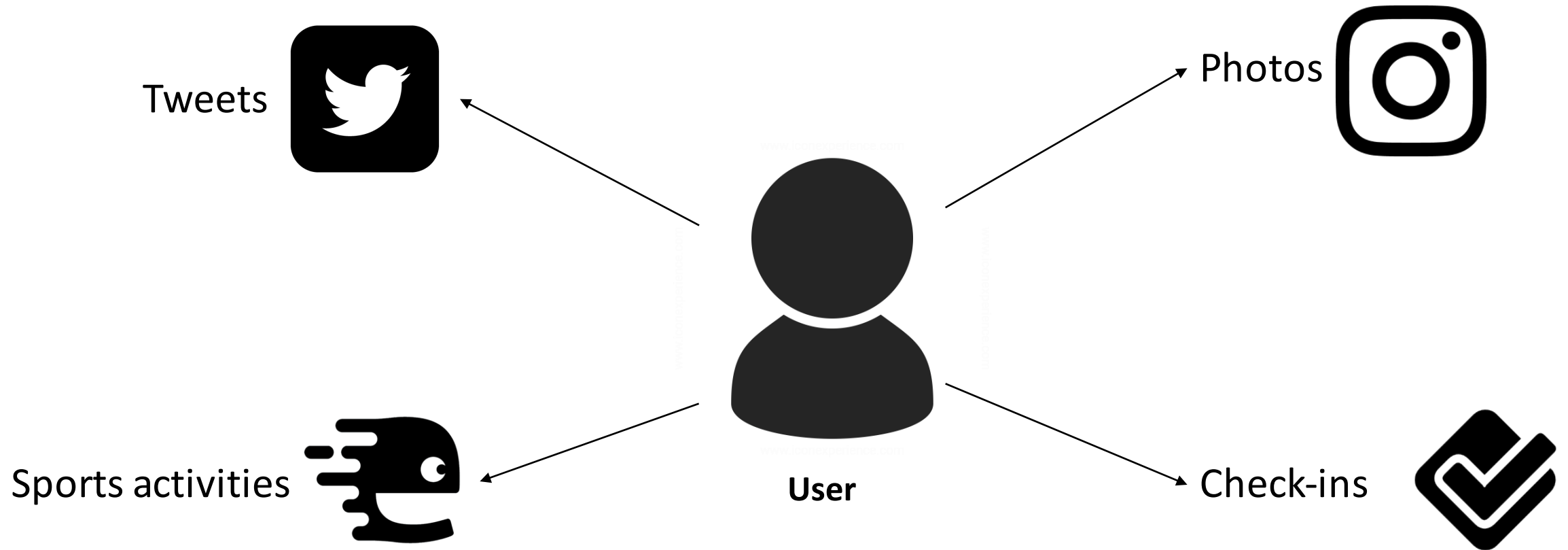
Towards User Personality Profiling from Multiple Social Media

Kseniya Buraya, Aleksandr Farseev, Andrey Filchenkov, Tat-Seng Chua

What is user personality?



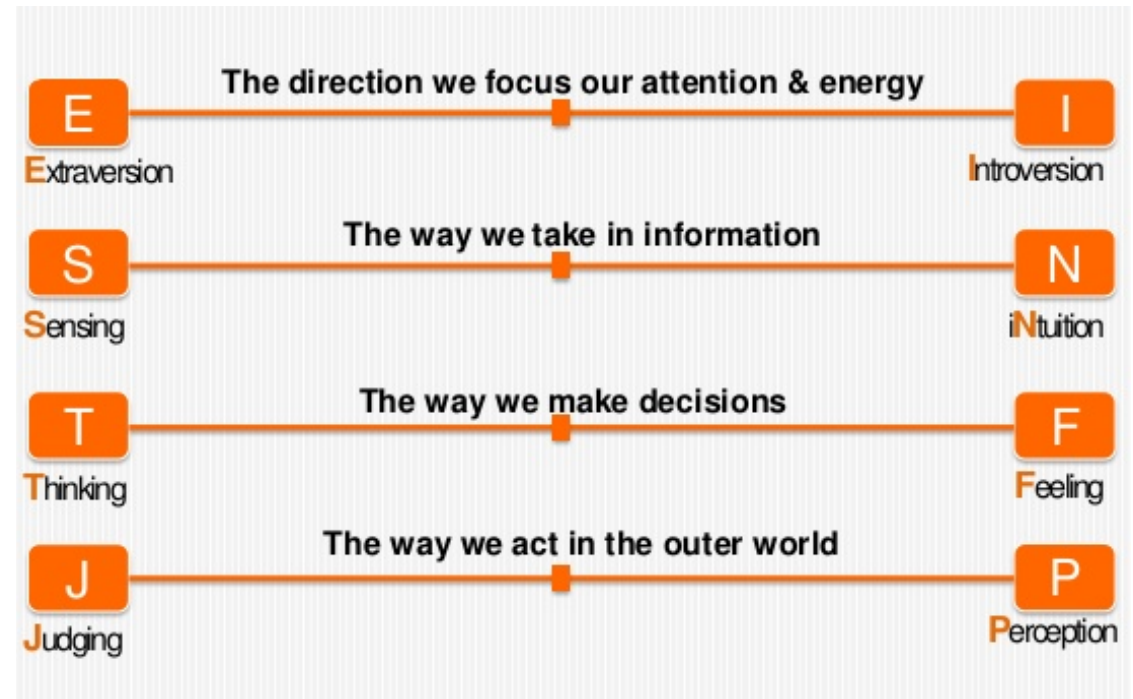
Social media and personality



Related works

Mayers-Briggs Type Indicator (MBTI)

MBTI – the typology, which is designed to exhibit psychological preferences in how people perceive the world around them and distinguishes **16 personality types**.



Why MBTI?



Social media services exceedingly affect and reflect the way how their users communicate with the world and among themselves.



MBTI categorization schema naturally fits social media research.

Challenges and goals

Challenges:

- Cross-source user identification
- Ground-truth collection
- Temporal changes of users' personality
- Data source fusion



Research Question:

Is it possible to boost personality performance by incorporating multi-source data?

Data

NUS-MSS – the largest available multi-source cross-region dataset.

It contains:

- **Foursquare** as a location data source;
- **Twitter** as a textual data source;
- **Instagram** as a visual data source;
- **Facebook** as a demographics-related ground truth source.

City	#users	#tweets	#check-ins	#images
Singapore	7,023	11,732,489	366,268	263,530
London	5,503	2,973,162	127,276	65,088
New-York	7,957	5,263,630	304,493	230,752

Ground Truth: Age, Education, Employment, Demography, Location, Relationship Status.

Features

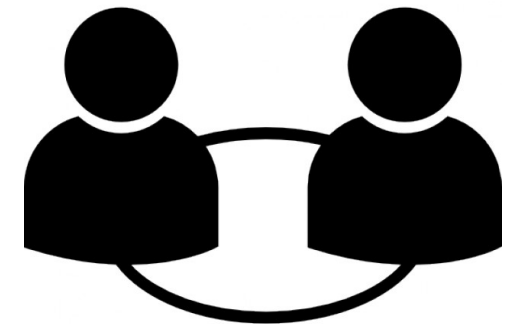
- **Location Features.** Data contains 592 venue categories from Foursquare.
- **Text Features:**
 - **LDA-based Features.** All documents were projected into a latent topic space using Latent Dirichlet Allocation (LDA) [2]. $T = 50$ topics.
 - **Linguistic Features.** For each user tweet list dataset contains 71 LIWC features.
 - **Heuristically-inferred Features.** The number of URLs, hashtags and user mentions, slang words, emotion words, emoticons; average sentiment score; linguistic style features.
- **Visual Features.** Each Instagram photo is mapped to the pre-defined image concept dictionary, comprising of 1000 daily life concepts.

Experiment

Based on the previous personality studies:

- Relationship status is closely related to human' personality;
- One's personality is often affected by his/her current relationship status;
- One's relationship status often depends on life partners' personality types match;
- The relationship status attribute can also be considered as a dynamic personal attribute, since it often changes over time, similarly to personality type

NUS-MSS Dataset contains information about users' relationship status.



Experiment

Because of relationship status is closely associated with one's personality, we selected it for evaluation in our study.

Step 1. Divide NUS-MSS dataset users into 'single' and 'not single'



Step 2. Fit Support Vector Machine, Logistic Regression and Gradient Boosting Decision Trees classifiers



Results

	Singapore	New York	London
Twitter	0.688	0.580	0.637
Instagram	0.677	0.540	0.613
Foursquare	0.666	0.800	0.590
Twitter, Instagram	0.865	0.851	0.655
Twitter, Foursquare	0.790	0.714	0.808
Instagram, Foursquare	0.780	0.710	0.714
Twitter, Instagram, Foursquare	<i>0.780</i>	<i>0.714</i>	<i>0.714</i>

Quality of relationship status profiling in terms of average accuracy

Results explanation

Future Work